

BAKALÁŘSKÁ PRÁCE

**Systém pro automatickou
registraci do katalogů**

System for Catalogue Registration

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

Zadání bakalářské práce

Student: **Jiří Kopel**
Studijní program: B2647 Informační a komunikační technologie
Studijní obor: 2612R025 Informatika a výpočetní technika
Téma: **Systém pro automatickou registraci do katalogů**
System for Catalogue Registration

Zásady pro vypracování:

Cílem práce je provést analýzu, návrh a implementaci desktopové aplikace, která bude řešit automatizaci procesu registrace do katalogů, vyhledávačů a redakčních systémů pro PR články. Zaměřte se na problematiku automatizované a poloautomatizované registrace do internetových katalogů a redakčních systémů.

1. Aplikace bude administrovat jednotlivé reklamní a zpětné odkazy.
2. Aplikace bude nabízet různé varianty registrovaného obsahu.
3. Aplikace bude schopna rozpoznat a řešit základní bezpečnostní prvky registrací.
4. Aplikace umožní přístup k registračnímu emailu klienta, tak aby mohlo docházet k automatickému potvrzování kontrolních registračních emailů.
5. Aplikace umožní budování farmy zpětných odkazů pro rychlé umístění na své stránky.
6. Aplikace vyhodnotí dokončení registrace, tedy jestli byl daný odkaz zařazen do katalogu.
7. Výsledné řešení testujte na sadě reálných projektů.
8. Definujte časové úspory při automatizaci registrace.

Seznam doporučené odborné literatury:

Podle pokynů vedoucího bakalářské práce.

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

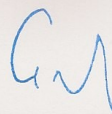
Vedoucí bakalářské práce: **Ing. Radoslav Fasuga, Ph.D.**

Datum zadání: 16.11.2012

Datum odevzdání: 07.05.2014



doc. Dr. Ing. Eduard Sojka
vedoucí katedry



prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prohlašuji, že jsem tuto bakalářskou/diplomovou práci vypracoval samostatně.
Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě : 17. 7. 2014

.....Kopel.....

Jiří Kopel

Abstrakt

Hlavním cílem této bakalářské práce je automatizace procesu registrace do internetových katalogů a vyhledávačů. Účelem je jednoduché a rychlé zlepšení SEO. Práce se dělí na dvě hlavní části. První část se zaměřuje na teoretickou stránku. Do toho patří seznámení s internetovými katalogy a vyhledávači a stručný popis hodnocení webového projektu vyhledávačem. Dále popisuje obecný postup při registraci do katalogu a rozdíly při vytváření automatizovaného registrátoru. Druhá část obsahuje návrh a popis implementace systému v .NET Frameworku, spolu s popisem funkcí a shrnutím dosažených výsledků.

Klíčová slova: automatizovaná registrace, katalog, vyhledávač, SEO

Abstract

The main goal of this bachelor thesis is the automation of registration process into internet catalogues and search engines. The purpose is simple and rapid improvement of SEO. The work is divided into two main parts. The first part focuses on theoretical side. This includes familiarization with online catalogs and search engines and a brief description of the project evaluation by web search engine. It also describes the general procedure for registration in the catalog and the differences in creating automated registrar. The second part contains a description of the design and implementation of the system in the .NET Framework, along with a description of features and a summary of achieved results.

Keywords: registration, catalogue, search engine, SEO

Seznam použitých zkratek a symbolů

API Application Programming Interface - rozhraní pro programování aplikací

HASH Jednoznačný identifikátor generovaný hašovací funkcí

HTML HyperText Markup Language

HTTP HyperText Transfer Protocol

JSON JavaScript Object Notation - způsob zápisu dat nezávislý na počítačové platformě

MIME Multipurpose Internet Mail Extensions - víceúčelová rozšíření internetové pošty

MVC Model View Controller - softwarová architektura rozdělující datový model aplikace do tří nezávislých komponent

OCR Optical Character Recognition - optické rozpoznávání znaků

SŘBD Systém řízení báze dat

URI Uniform Resource Identifier - jednotný identifikátor zdroje

XML Extensible Markup Language - rozšiřitelný značkovací jazyk

Obsah

1.	Úvod	1
2.	Existující řešení	2
2.1.	RegTool2	2
2.2.	SEO PowerSuite – LinkAssistant	3
3.	Optimalizace pro vyhledávače	5
3.1.	Co je to vyhledávač	5
3.2.	Katalog odkazů	5
3.3.	Proč registrovat do vyhledávačů	6
3.4.	Proč registrovat do katalogů	6
3.5.	Možnosti penalizace	7
4.	Postup přidávání stránek do katalogů	8
4.1.	Průběh registrace	8
4.2.	Rozdíly v postupu registrace	9
5.	Analýza informačního systému	15
5.1.	Specifikace požadavků na IS	16
5.2.	Datová analýza	16
5.3.	Funkční analýza	18
6.	Implementace aplikace	20

6.1.	Run-time prostředí	20
6.2.	Zvolený SŘBD	20
6.3.	Registrační část.....	21
6.4.	Emailový klient.....	23
6.5.	Administrační aplikace	25
6.6.	Aktualizace aplikace ze serveru	26
7.	Využité technologie a nástroje	29
7.1.	Microsoft Visual Studio 2012.....	29
7.2.	LINQ.....	29
7.3.	Json	30
8.	Závěr	31
9.	Citovaná literatura	32

Seznam obrázků

Obrázek 1 – RegTool2	4
Obrázek 2 - LinkAssistant	4
Obrázek 3 – Principiální postup registrace do katalogu	8
Obrázek 4 – Captcha příklady	12
Obrázek 5 – Work-flow diagram registrace s nutností vytvoření uživatelského účtu	14
Obrázek 6 - Use-case diagram	15
Obrázek 7 - Data flow diagram nulté úrovně - kontextový diagram	18
Obrázek 8 - Diagram případů užití	19
Obrázek 9 – Diagram postupu registrace.....	21
Obrázek 10 – Hlavní obrazovka s výběrem projektu	22
Obrázek 11 – Automaticky vyplněné prvky v katalogu	23
Obrázek 12 – Administrační program ve stádiu provazování formulářového pole s databází	26
Obrázek 13- Postup aktualizace dat ze serveru	27
Obrázek 14 – API přenosu dat mezi serverem a aplikací	28

Seznam kódů

Zdrojový kód 1 - Příklad hlavičky emailové zprávy	24
---	----

1. Úvod

Cílem této bakalářské práce je prozkoumání postupů a možností pro automatické a poloautomatické registrace do katalogů a vyhledávačů na internetu. Na základě zjištěných informací pak vytvořit poté vytvořit systém, jenž by zcela či částečně zautomatizuje registraci do těchto katalogů.

Důvodů proč provádět tyto registrace je několik a budou dále rozebrány v této práci, ale základní myšlenkou je konečné zvýšení návštěvnosti stránek registrovaných do katalogů. Metod jak tohoto cíle dosáhnout je několik.

Tato práce je zaměřena převážně na metodu využívající registrace do katalogů, proč a kdy je vhodné využít této metody a v jakých případech může mít využití této metody dokonce i opačný efekt. Dále obsahuje popisy způsobů registrace do katalogů a možnosti jak se některé katalogy brání proti automatizovanému vkládání dat. Jsou zde také uvedena možná řešení na tyto pokusy o zabránění automatickému vkládání dat do katalogů.

Praktická část této práce obsahuje analýzu informačního systému a datovou analýzu. Následuje popis implementace, jenž obsahuje popis použitých technologií a vlastní implementaci klíčových funkcí systému.

2. Existující řešení

2.1. RegTool2

Jedná se o aplikaci pro poloautomatickou registraci projektů do katalogů. Aplikace je placená a obsahuje aktualizovaný seznam katalogů. Každý uživatel má data uložena online a tudíž přístupná odkudkoliv a stále zálohována. Každý projekt lze, pro větší přehlednost, označit vlastními štítky nebo lze projekty filtrovat podle data vytvoření, url adresy nebo registračního emailu. Aplikace nabízí, pro zvýšení bezpečnosti před ztrátou, zálohu dat do souboru na počítač.

Při zakládání nového projektu je zapotřebí zadat webovou adresu, titulky webu a krátký popis a registrační email, na který budou většinou přicházet informační zprávy z katalogů (potvrzení o přidání, apod.). Dále je možno přidat další titulky webu a stručné popisy, podrobnější popisy, klíčová slova, obrázek jako náhled webu či jiné vlastní texty pro rychlé přidávání. K registračnímu emailu lze přidat registrační jméno a heslo pro případ potřeby registrace na stránkách katalogu. Pro katalogy firem a eshopů lze přidat informace o firmě jako název, adresa, IČO, DIČ, telefon či firemní email.

Po výběru projektu poté vybíráme katalogy, do kterých chceme provést. Katalogy lze třídit a filtrovat podle typu (vyhledávače, katalog firem, katalog eshopů, atd.), typu registrace, jazyku, PageRanku, Seznam Ranku. Lze vytvořit vlastní balíčky katalogů, pro případ kdy budeme přidávat několik projektů do stejných katalogů.

Samotné registrace projektů probíhá otevřením webových stránek katalogu v okně aplikace a některých případech pomocí navigace uživatele vyhledáním registračního formuláře na stránkách katalogu. Po načtení stránky s registračním formulářem dojde k automatickému vyplnění položek, jejichž informace uživatel zadal při vytváření projektu. V případě potřeby uživatel doplní další informace (buď ručně, nebo pomocí menu vybere informace zadané při vytváření projektu) a pokračuje v registraci.

Hlavní výhodou aplikace je obsáhlý a aktualizovaný seznam katalogů a automatické vyplnění opakovaných informací jako adresa projektu, popis, apod., čímž sníží čas potřebný pro vyhledání katalogů a vyplnění stále se opakujících se informací.

Nevýhodou je potřeba navigace skrz stránky katalogu, jenž je patrné již ze zaměření na poloautomatickou registraci. Tato možnost sice zvyšuje rychlost oproti klasickému vkládání dat ručně, ale například u jednodušších katalogů se může jednat o zbytečné zpomalení.

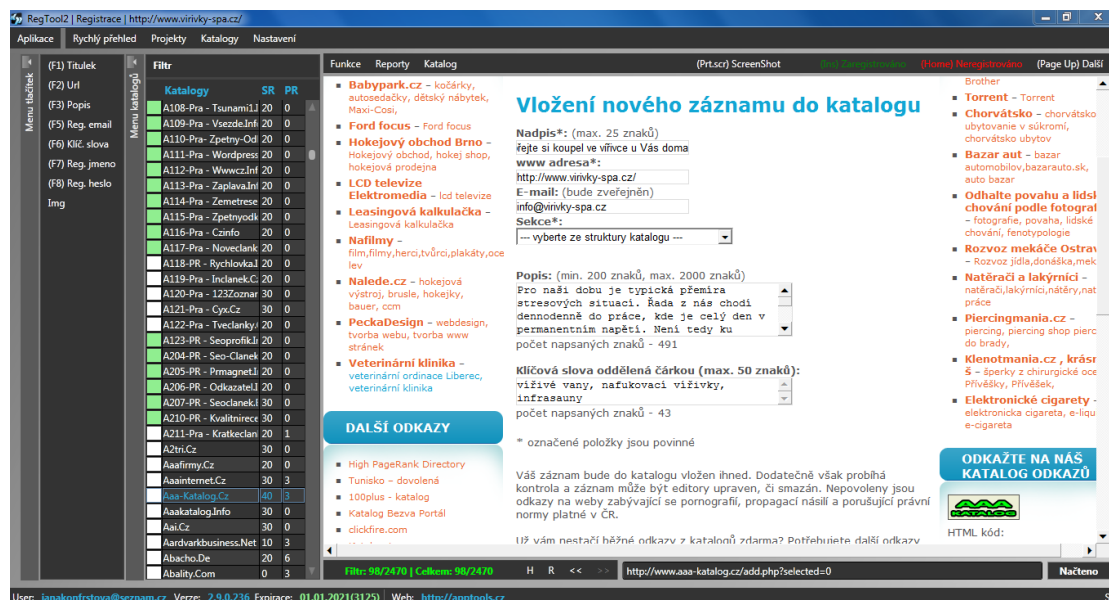
2.2. SEO PowerSuite – LinkAssistant

LinkAssistant je profesionální nástroj pro registraci do katalogů stránek, jenž je součástí balíčku SEO Power Suite. Tento balíček obsahuje dále aplikace na zjištění odkazové strategie konkurenčních stránek, sledování vývoje ranku vašich stránek i stránek konkurence a nástroj pro analýzu webových stránek. Nástroj LinkAssistant se zaměřuje na pomoc při budování odkazů nejen přes katalogy ale i pomocí webových partnerů.

Součástí tohoto nástroje je plnohodnotný emailový klient, který je rozšířený o automatické funkce, které zjednodušují správu zpráv od katalogů a webových partnerů.

Základní funkcí nástroje je vyhledání a správa vhodných webových partnerů. K vyhledání vhodných partnerů aplikace využívá několika způsobů, např. pomocí klíčových slov, která zadá uživatel nebo která aplikace vygeneruje z webové stránky. Dále umožňuje prohledat web na stránky odkazující na konkurenci. Po vyhledání mnoha možných partnerů lze tento seznam omezit stanovením hodnoty jednotlivých partnerů pomocí několika faktorů.

Dále aplikace obsahuje funkci kontroly, zda dané stránky obsahují zpětné odkazy na naše stránky. Součástí této funkce je i detekce, zda kontrolované stránky nepoužívají některou z funkcí, která by znemožnila vyhledávačům tyto odkazy nalézt a kontroluje jaká je důležitost těchto stránek podle vyhledávačů. Mezi zajímavé možnosti patří i tvorba víceúčelových reportů. Umožňuje výběr mezi několika typy reportů pro různé cíle, od analýzy průběhu tvorby



Obrázek 1 – RegTool2

Partners							
Partners							
#	Include in Link Direc...	Partner Domain	Total links (...)	Entrance Date	Verification Passed	Backlink Page	Domain PR (Partner)
185	✓	http://lberdeen.ironbirds.mlb.c...	541	Aug 4, 2011	Yes	http://lberdeen.ir...	PR: 5
186	✓	http://indianapolis.indians.mlb...	541	Aug 4, 2011	Yes	http://indianapol...	PR: 5
187	✓	http://www.essexfamilydental.c...	39	Aug 4, 2011	Yes	http://www.essex...	PR: 1
188	✓	http://www.gstnh.com/granites...	22	Aug 4, 2011	Yes	http://www.gstnh...	PR: 3
189	✓	http://www.geomstevensinsura...	71	Aug 5, 2011	Yes	http://www.geom...	PR: 2
190	✓	http://lexquisite-productions.com	11	Aug 5, 2011	Yes	http://lexquisite-p...	PR: 2
191	✓	http://www.bangorsymphony.org	48	Aug 5, 2011	Yes	http://www.bang...	PR: 4
192	✓	http://www.wolfenews.com	54	Aug 5, 2011	Yes	http://www.wolfe...	PR: 3
193	✓	http://inhada.com	38	Aug 5, 2011	Yes	http://inhada.com...	PR: 5
194	✓	http://inhdsr.org	32	Aug 5, 2011	Yes	http://inhdsr.org/...	PR: 4
195	✓	http://www.forestsofociety.org	31	Aug 5, 2011	Yes	http://www.forest...	PR: 5
196	✓	http://www.healthymh.com	47	Aug 8, 2011	Yes	http://www.health...	PR: 5
197	✓	http://atfscmetrus.org	0	Aug 8, 2011	Yes	http://atfscmetrus...	PR: 0
198	✓	http://www.gath-nh.org	36	Jul 28, 2011	Yes (Wrong anchor)	http://www.gath...	PR: 4
199	✓	http://www.wgplm.com	49	Jul 28, 2011	Yes (Wrong anchor)	http://www.wgpl...	PR: 4
200	✓	http://www.nh1.edu	94	Jul 28, 2011	Yes (Wrong anchor)	http://www.nh1.e...	PR: 5
201	✓	http://igsqc.com	217	Jul 28, 2011	Yes (Wrong anchor)	http://igsqc.com/c...	PR: 5
202	✓	http://www.mainehr.com	21	Jul 28, 2011	Yes (Wrong anchor)	http://www.maine...	PR: 3
203	✓	http://www.nh1gc.org	92	Jul 28, 2011	Yes (Wrong anchor)	http://www.nh1gc...	PR: 5
204	✓	http://www.wopular.com	358	Jul 28, 2011	Yes (Wrong anchor)	http://www.wopul...	PR: 5
205	✓	http://www.gshrcconference.org	55	Jul 28, 2011	Yes (Wrong anchor)	http://www.gshrc...	PR: 4
206	✓	http://www.walkwithme.org	30	Jul 29, 2011	Yes (Wrong anchor)	http://www.walk...	PR: 5
207	✓	http://inemestic.com	373	Jul 29, 2011	Yes (Wrong anchor)	http://inemestic.c...	PR: 5
208	✓	http://humanresources.vermont...	75	Jul 29, 2011	Yes (Wrong anchor)	http://humanres...	PR: 5

Obrázek 2 - LinkAssistant

3. Optimalizace pro vyhledávače

Do tzv. optimalizace pro vyhledávače (Search Engine Optimization = SEO) lze zahrnout rozmanitou skupinu aktivit pro zvýšení cíleného provozu, který přichází na daný webový projekt z vyhledávacích strojů. [1] V případě systému pro registraci do katalogů se jedná o zvyšování počtu zpětných odkaz (tzv. linkbuilding). Toto patří mezi off-page faktory SEO - faktory, které nelze ovlivnit na úrovni webových stránek.

3.1. Co je to vyhledávač

Vyhledávač (internetový vyhledávač, vyhledávací stroj apod.) je internetová služba procházející webové stránky po celém internetu. Pomocí hypertextových odkazů na stránkách se dostává na další podstránky nebo jiné weby a indexuje jejich obsah. Když pak uživatel zadá do vyhledávače fráze pro vyhledání v internetu, vyhledávač vrátí webové stránky, které nejvhodněji odpovídají hledaným frázím. Příklady internetových vyhledávačů jsou světově známý vyhledávač Google, Yahoo!, Bing nebo český Seznam.

3.2. Katalog odkazů

Jedná se o webové stránky, obsahující databázi URL adres. Tyto adresy odkazují na webové projekty a tím zvyšují počet zpětných odkazů, vedoucích k danému webovému projektu.

Některé katalogy nabízejí placenou a neplacenou variantu registrace. U neplacené registrace se často zpětný odkaz v katalogu objeví až po určitém čase. Naopak placená registrace většinou nabízí brzký výskyt zpětného odkazu v katalogu a mnohdy i registraci do dalších katalogů.

Dále se vyskytují katalogy s možností výměny odkazů. V praxi to znamená, že odkaz na registrovaný webový projekt se v katalogu objeví až poté, co se odkaz na katalog objeví na webových stránkách projektu. Alternativou jsou katalogy, u kterých výměna odkazu není nutná, ale pokud je provedena, tak poté tento webový projekt je nějakým způsobem zvýhodněn např. zobrazením odkazu na titulní stránce katalogu.

3.3. Proč registrovat do vyhledávačů

Registraci do vyhledávače sice vždy nezíská webový projekt další zpětný odkaz, ale dříve se objeví ve výsledcích vyhledávání daného vyhledávače. Pokud registraci do vyhledávače neprovedete sami, musíte se spolehnout na to, až webový projekt najde robot (neboli crawler či spider) daného vyhledávače. Pokud nemá webový projekt zpětný odkaz, umístěný na silném webu, pak první indexování robotem a tím i první výskyt ve výsledcích vyhledávání může trvat i několik týdnů či měsíců. (1)[1]

3.4. Proč registrovat do katalogů

Jedním z kritérií ovlivňující pozici webového projektu ve výsledcích vyhledávání internetového vyhledávače je množství a kvalita zpětných odkazů. Kvalita zpětného odkazu závisí na hodnocení relevantnosti webové stránky (katalogu). Hodnocení provádí vyhledávač a je označováno jako „page rank“ u Google, „srank“ u Seznamu apod. Využitím velkého množství slabých odkazů lze částečně doplnit nebo vyvážit silný zpětný odkaz. Vždy je však výhodnější získání silných zpětných odkazů.

Mnoho lidí si myslí, že čím vyšší bude mít ranky, tím se úměrně zvýší jejich pozice ve vyhledávačích. Tato domněnka, ale není úplně správná. Můžete si udělat jednoduchý test: vyhledejte si např. na Googlu nějaké slovo – klidně může být na 10. místě prezentace s PR 5 a první může být web s PR 3. [3] Katalogy, na rozdíl od vyhledávačů, nepoužívají automatické procházení internetu pomocí robotů. Přidávání odkazů je tedy závislé pouze na invenci ze strany provozovatele webového projektu. [1] Obvykle pak administrátor katalogu ručně kontroluje registrované webové projekty.

Podle některých zdrojů mohou zpětné odkazy být rozhodujícím faktorem při určování pozice webového projektu ve výsledcích vyhledávání.

O síle zpětných odkazů vypovídají tzv. Google bomby. Jedná se v podstatě o soustředěnou snahu odkázat na danou stránku na konkrétní fráze nebo klíčové slovo, a to především takové, které se na dané stránce vůbec nenachází. V naprosté většině k tomu dochází bez vědomí majitele nebo provozovatele stránek. [2] Proti takovým

praktikám se však vyhledávače snaží bránit. Přestože to může být dobrá reklama, nicméně to snižuje důvěryhodnost vyhledávače jako takového.

3.5. Možnosti penalizace

Je důležité vědět, že náhlé zvýšení počtu zpětných odkazů může být vyhledávači penalizováno. Vyhledávač totiž primárně počítal s tím, že spokojení uživatelé webového projektu umístí odkaz na svůj blog, webové stránky apod. a tímto způsobem budou pozvolna vytvářeny zpětné odkazy. Velké množství zpětných odkazů by tedy mělo naznačovat velkou oblibu webového projektu. Tím pádem vyhledávač ve výsledcích upřednostňuje takto oblíbený projekt oproti podobným webům. Vytváření katalogových odkazů ale provádí provozovatel webového projektu, snažící se získat co nejlepší pozici ve vyhledávání co nejdříve. Další možnost, jak získat velké množství zpětných odkazů v krátkém čase, je založení několika fiktivních webů obsahující příslušné zpětné odkazy. Toto ovšem je vyhledávači striktně označováno jako nekalá praktika.

Tvorba dobrého textového odkazu je něco, co se ne každý naučí. Proto můžete chtít sáhnout po automatických systémech tvořících obsah a stránku takovým generickým obsahem naplnit, abyste získali velké množství obsahu a dosáhli lepšího hodnocení stránky. Problémem ale je, že vyhledávač dokáže tento duplikovaný obsah odhalit a stránky penalizovat.

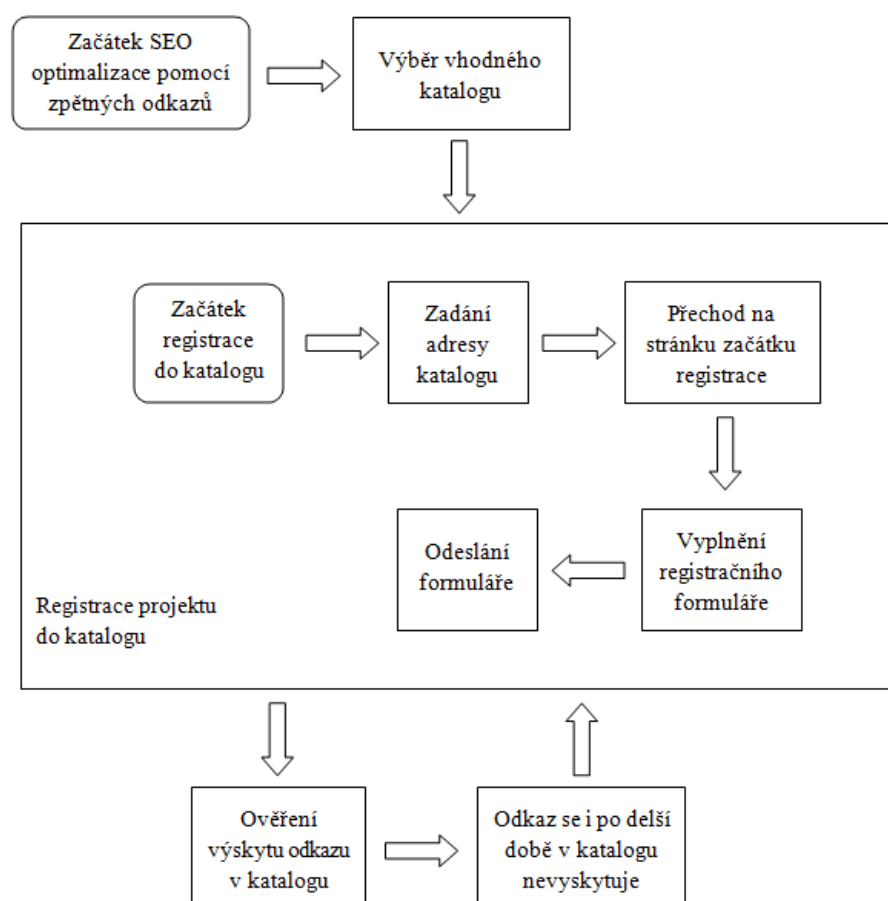
Pokud chceme rychle zlepšit pozici ve výsledcích vyhledávání zpětných odkazů a zároveň se vyhnout penalizacím, pak je nejvhodnější použití pouze „silných“ katalogů. Jedná se o katalogy, které mají u vyhledávačů kvalitní hodnocení. Tyto katalogy ovšem bývají placené, nebo alespoň vyžadují výměnu odkazu.

4. Postup přidávání stránek do katalogů

Postup, podle kterého se provádí registrace do katalogů, je v zásadě jednotný. Protože existuje několik typů katalogů, jsou jednotlivé kroky rozdílné v jednotlivých detailech. To nám nezabraňuje ve vytvoření schématu, podle kterého se můžeme řídit ve velké většině případů.

4.1. Průběh registrace

Níže uvedený diagram nám znázorňuje zjednodušený průběh procesu registrace projektu do katalogu. I když se jedná o zjednodušený postup, v mnoha případech se dá úspěšně využít, a to i s velmi minimálními až žádnými úpravami.



Obrázek 3 – Principiální postup registrace do katalogu

Nejdůležitějším a nejvíce měněným krokem je „Vyplnění registračního formuláře“. Tato část, v závislosti na katalogu, může obsahovat jeden až několik kroků. V zásadě se jedná o přečtení webového formuláře a rozpoznání, jaké informace je potřeba vyplnit. Jedná se o krok, který uživatel provádí bez toho, že by se nad tím zamýšlel.

Při automatizovaném vyplňování ovšem takovému instinktivní rozhodování je velmi obtížné. Proto je potřeba vytvořit jakousi databázi metadat, podle kterých registrační systém rozhodne kam má jaká data vložit. Nejjednodušším způsobem je vytvoření párů *identifikátor_prvku: typ_údaje*. Ne vždy je ovšem takovýto způsob vhodný nebo možný.

4.2. Rozdíly v postupu registrace

Jak jsme si ukázali výše, pro úspěšnou registraci je potřeba provést danou posloupnost kroků. Principiálně se postup při registraci nemění, ovšem realizace se vždy mění v mnoha ohledech. Bohužel každý katalog (v tomto případě budeme brát katalogové enginy jako jeden katalog) je vždy unikátní a v jistých směrech se odlišuje od základního postupu. Většina katalogů často obsahuje nějaký typ zabezpečení proti automatizovaným strojům.

- **Rozdělení zadávání údajů do kroků**

Nejjednodušší a asi nejčastější rozdíl od základní posloupnosti kroků je rozdělení vyplňování registračního formuláře do několika kroků. Jedná se o jednoduchý způsob, jak velkým množstvím požadovaných dat neodradit uživatele od registrace.

Toto ovšem přináší první obtížnosti při tvorbě systému pro automatickou registraci. Katalog totiž musí nějakým způsobem vědět, jaká data byla vyplněna v předchozích krocích. Mezi nejzákladnější způsoby patří ukládání dat do skrytých polí formuláře (více odrážka níže) a uložení dat v session na serveru.

Práce s daty ve skrytých formulářích je velmi jednoduchá. Může nám umožnit přeskočení jednotlivých kroků přímo na poslední. V tomto kroku nám pak jen stačí vyplnit data do těchto polí a odeslat celý formulář.

V případě, kdy katalog používá session, nelze něčeho takového použít. Data jsou uložena na serveru, a tudíž musíme jednotlivé kroky provést v daném sledu. Dalším problémem je potřeba přenášení identifikátoru klienta na server. To se většinou provádí použitím cookies, která klient odesílá spolu s HTTP požadavkem. Výsledkem je tedy nutná možnost použití HTTP cookies v registračním systému.

- **Výběr kategorie**

I když se to tak na první pohled nemusí zdát, tak výběr kategorie může být velmi obtížná část automatické registrace do katalogu. Je to tím, že jednotlivé katalogy se liší v počtu a názvech kategorií a v tom jestli mají podkategorie. Následný výběr kategorie pak může být velmi náročný a mnohdy bez pomoci uživatele až nemožný (např. při velkém množství podkategorií).

Z technického hlediska výběru kategorie pak existuje několik způsobu, jak je možno ve formuláři vybrat kategorii.

- formulářový prvek SELECT – jedná se o velmi jednoduchý a často používaný způsob zobrazení a výběru kategorie. Při menším množství kategorií je velmi přehledný a rychlý. Z hlediska automatické registrace se jedná o snadno použitelný prvek formuláře.
- dva prvky SELECT – používaný ve spojení s podkategoriemi. První prvek slouží k výběru kategorie, do druhého se pak pomocí JavaScript kódu nějakým způsobem vloží položky podkategorie. Registrační systém v takovémto případě musí buď obsahovat předpřipravený seznam možných položek, nebo musí umět zpracovat změny v druhém prvku SELECT.
- Přepínače – ve spojení s výběrem kategorií se jedná o nepříliš používaný formulářový prvek. Je to z důvodu, že pro tento prvek je potřeba zobrazovat všechny možnosti přímo na stránce oproti SELECT prvku. Lze se s tímto prvkem setkat při zadávání jiných údajů do katalogu. Výběr možnosti je snadný.
- textové pole – je většinou doplněné o našeptávač, jenž vypíše možné kategorie podle zadaného textu. Nejvhodnější způsob automatického vyplnění je pomocí předpřipraveného seznamu kategorií.

- registrace do právě procházené kategorie odkazů – některé katalogy požadují jako první krok výběr kategorie. Při automatizované registraci je tedy také potřeba nejdříve znát, do které kategorie daný projekt patří. Až poté je možno vkládat ostatní data. Při tomto způsobu je zase nejvhodnější mít předpřipravený seznam kategorií, který navíc obsahuje odkaz na stránky obsahující formulář pro přidání projektu do dané kategorie.

- **Skrytá pole**

Formuláře běžně obsahují skrytá pole, jejichž hodnota nelze za běžných okolností uživatelem změnit ani pro něj není viditelná. Tyto hodnoty využívá server, buď k rozlišení o který formulář se jedná, nebo jako základní prvek zabezpečení. Tato pole je potřeba nalézt a v případě, kdy lze zasílat vyhodnocovacímu skriptu data (zcela automatizovaný postup), si uchovávat jejich hodnotu, aby nebylo nutné opětovně načítat webovou stránku a tento údaj hledat. V případě předvyplňování není extrahování takové hodnoty nutné.

Některé formuláře naproti tomu generují hodnotu skrytého pole jako nějaký identifikátor relace s omezenou časovou platností jakožto zabezpečovací prvek. Protože musejí být tyto hodnoty vždy odeslány, tak v případě, kdy lze zasílat vyhodnocovacímu skriptu data (zcela automatizovaný postup), musí být načtena daná webová stránka a údaj extrahován z jejího zdrojového kódu. V případě předvyplňování není extrahování takové hodnoty nutné.

- **Záchytné obrázky**

Záchytné obrázky, neboli captcha, je způsob zabezpečení webových formulářů proti automatům. Základní myšlenkou je využití obrázku obsahující text (náhodné znaky nebo slova), který musí uživatel opsat do pole formuláře.

CAPTCHA je akronym pro „*completely automated public Turing test to tell computers and humans apart*“, tedy „plně automatický veřejný Turingův test k odlišení počítačů a lidí“. Test spočívá zpravidla v zobrazení obrázku s deformovaným textem, přičemž úkolem uživatele je zobrazený text opsat do příslušného vstupního políčka. Předpokládá se, že lidský mozek dokáže správně rozeznat i deformovaný text, ale

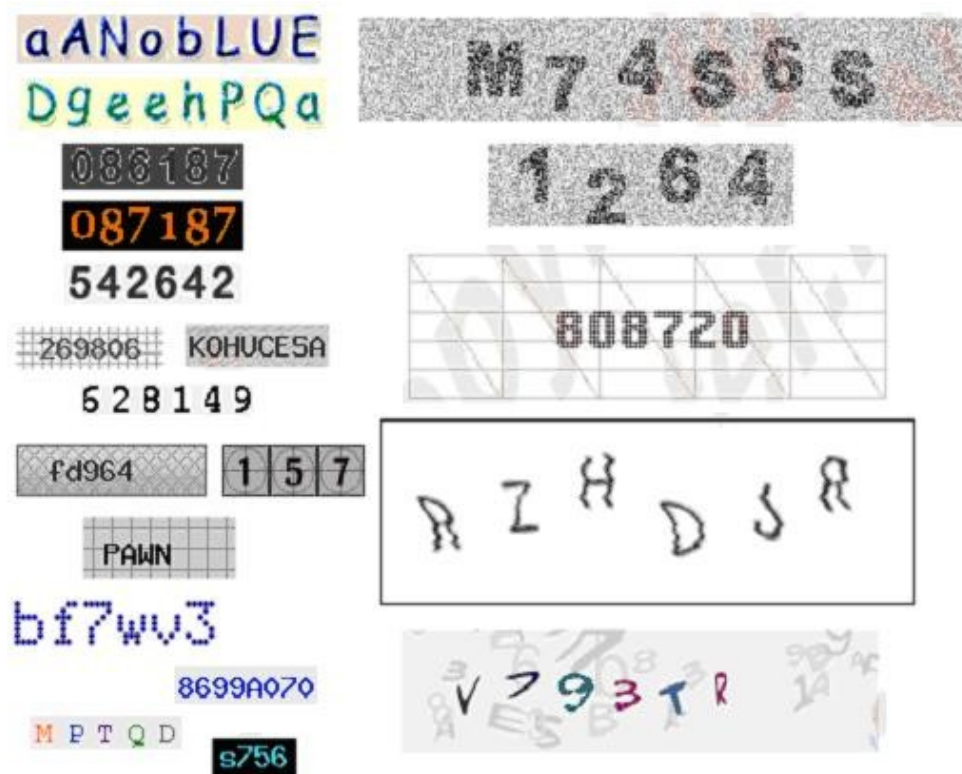
internetový robot při použití technologie OCR nebude schopen text správně rozpoznat.

(4)

Základní verzí je jednoduchý obrázek, obsahující jednobarevné pozadí a znaky jedné barvy. Tento typ je velmi jednoduchý na automatické zpracování, protože neobsahuje žádné rušivé elementy nebo deformace a tudíž lze jednotlivé znaky jednoduše identifikovat. Identifikace spočívá v rozpoznání znaku a následného porovnání se vzorovou databází.

Možností ke zvýšení obtížnosti automatického zpracování je několik. Základní, jako je vícebarevné pozadí a změna barev znaků, vede k potřebě složitějších algoritmů pro rozpoznání znaku od pozadí. Dále je zde prolínání jednotlivých znaků, které má za následek detekci více znaků jako jediný neznámý znak, nebo i jako jeden jiný znak (např. „ct“ jako „d“). Další oblíbenou formou je deformace textu. Hlavní způsoby jsou pomocí náhodného zvlnění a pootočení znaků.

Pro rozluštění takto obtížných záchytných obrázků je vhodnější využití uživatele. Tomu záchytný obrázek zobrazíme a on zapíše zobrazené znaky.



Obrázek 4 – Captcha příklady

- **Generování formuláře JavaScriptem**

Pokud se o vytvoření formuláře stará JavaScript, není možné ve zdrojovém kódu HTML stránky nalézt formulářové prvky. Pokud bychom i v takovémto případě požadovali zcela automatickou registraci, musel by být implementován stroj, interpretující jazyk JavaScript. Díky němu by bylo možné nalézt pole formuláře a extrahovat jejich názvy. Před postupem registrace by se muselo opět uvážit, jaké další odlišnosti od základního postupu tato registrace obsahuje.

- **Morfující názvy formulářových polí**

Formulářová pole dostanou po každém načtení či obnovení webové stránky s formulářem nový název a identifikátor není uveden vůbec. Poměrně málo používaná, ale velice nepříjemná technika zabezpečení. Znemožňuje využití seznamu názvů polí, uložených v DB. V takovémto případě je potřeba zjistit pozici pole a odkazovat se pomocí takového identifikátoru.

- **Kontrola referrera**

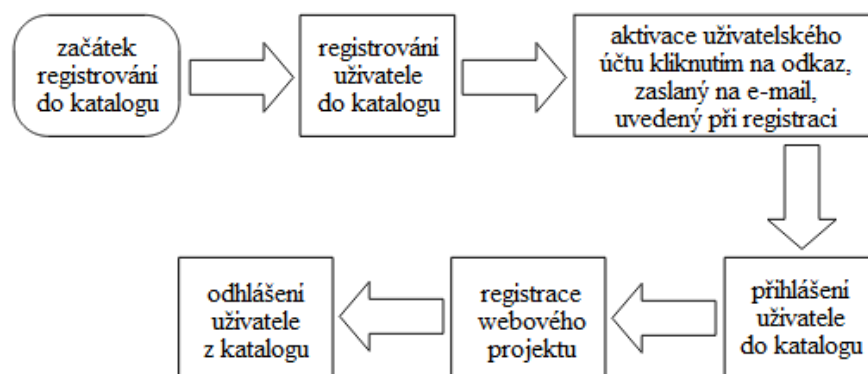
HTTP referrer je v informatice označení pro URI, ze kterého byla webová stránka navštívena. Údaj *HTTP referrer* zapisuje webový prohlížeč do hlavičky HTTP dotazu pro webový server, který ho může dále zpracovat. [5] Tato kontrola může mít několika využití. Od základního logování, přes sledování zdrojů návštěvnosti webových stránek až po využití v SEO.

Kontrola referrera patří mezi prvky ochrany proti sociálnímu inženýrství a tím pádem proti automatizovaným strojům. Díky tomu, že referrer je součástí HTTP hlavičky, se dá tato kontrola velice snadno obejít. Stačí jen před odesláním požadavku upravit hlavičku HTTP dotazu.

- **Nutnost vytvoření uživatelského účtu**

Další prvek zabezpečení katalogů, se kterým se lze setkat, je omezení v přidávání odkazů pouze zaregistrovaným a přihlášeným uživatelům. Registrování uživatele se provádí vyplněním formuláře, odlišného od toho pro registraci webového projektu. Po vyplnění a odeslání formuláře obvykle přichází na zadaný emailový účet

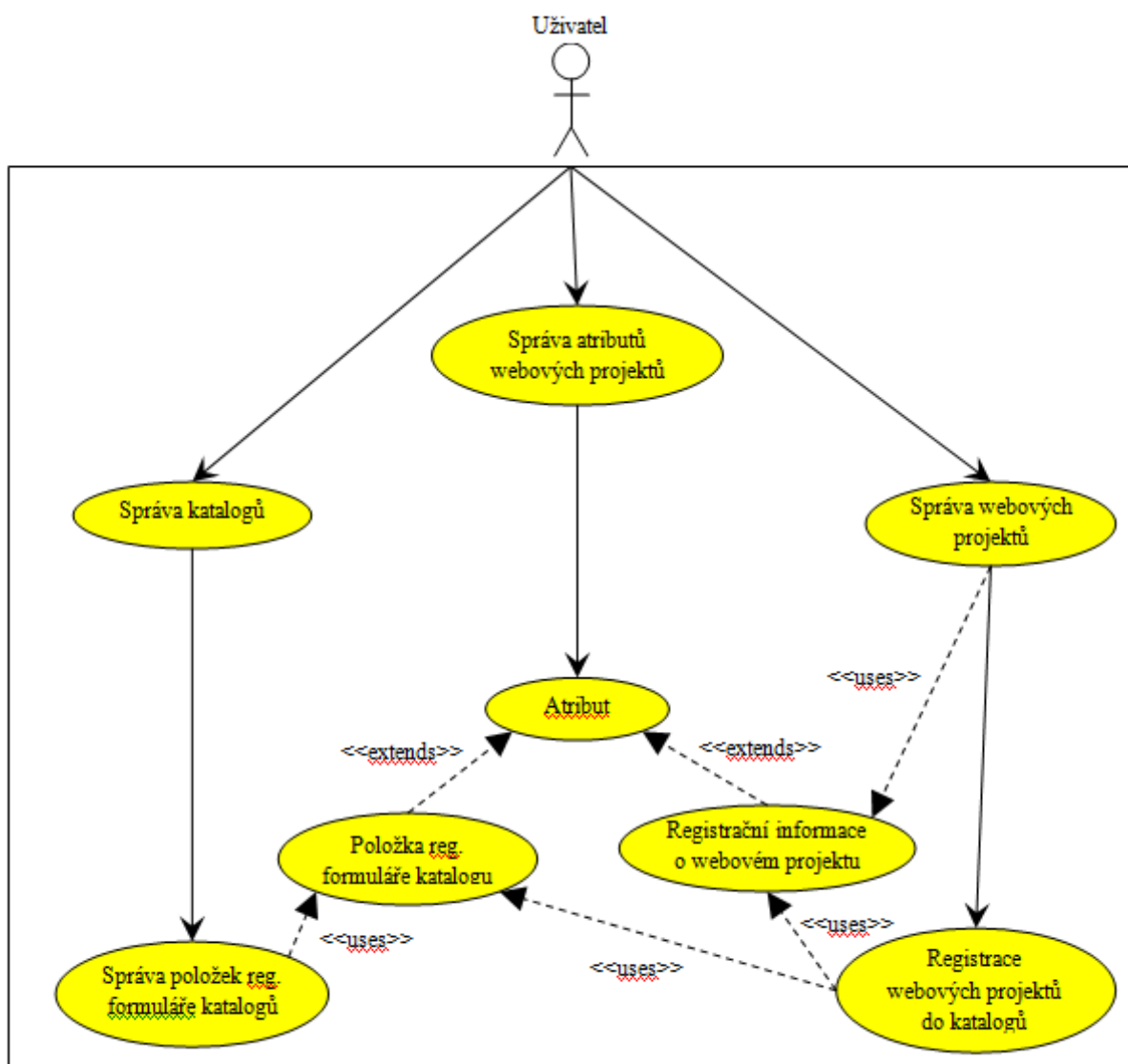
email s potvrzovacím odkazem. Po kliknutí na odkaz je účet uživatele aktivován a ten se již může přihlásit a registrovat webové projekty do katalogu. Jedná se o rozšíření postupu uvedeného na začátku kapitoly. Registrace do takovýchto katalogů vyžaduje systém, jenž umožňuje předávat informace o session, jelikož je potřeba udržet informace o uživateli. Upravené schéma postupu registrace znázorňuje diagram na obrázku níže.



Obrázek 5 – Work-flow diagram registrace s nutností vytvoření uživatelského účtu

5. Analýza informačního systému

Informační systém (IS) je soubor lidí, technologických prostředků a metod, které zabezpečují sběr, přenos, zpracování a uchování dat za účelem tvorby prezentace informací pro potřeby uživatelů. Informacemi míníme sdělení, které odstraňuje nejistotu nebo nevědomost, daty míníme jakékoli zaznamenané poznatky či fakta. Jako zvláštní pojem zde vystupuje také znalost představující zobecnění poznání určité části reality. Informaci je možno také chápat jako data s nějakým přidaným významem (data + význam). [4]



Obrázek 6 - Use-case diagram

5.1. Specifikace požadavků na IS

Informační systém budou využívat uživatelé, snažící se o zvýšení pozice ve výsledcích vyhledávačů. Budou potřebovat data o projektu, která budou muset zadat do systému a následně je pomocí systému registrovat do katalogů.

- **Funkční požadavky**

- Přehled a správa atributů webových projektů, požadovaných při registraci do katalogů
- Přehled a správa katalogů
- Přehled a správa webových projektů
- Registrace webového projektu do katalogu
- Ověření výskytu odkazu na webový projekt v katalogu

- **Vstupy IS**

- Obecné informace o katalogích typu webová adresa, jazyk, typ apod.
- Obecné informace o webových projektech typu webová adresa, název, jazyk apod.
- Informace o katalogích, potřebné pro vytvoření postupu registrace projektu
- Informace o webových projektech, které budou použity jako údaje při registraci do katalogů

- **Výstupy IS**

- Přehled webových projektů
- Přehled atributů webových projektů, požadovaných při registraci do katalogů
- Přehled katalogů a vyhledávačů
- Přehled položek registračního formuláře katalogu
- Výsledek automatické registrace do katalogu
- Předvyplněný registrační formulář

5.2. Datová analýza

Databáze patří mezi nepostradatelné komponenty systému. Uchovávají se v ní všechna perzistentní data a struktura jejich uspořádání musí odrážet vazby mezi objekty (entitami) v realitě. Návrh její struktury má vliv na odezvu systému. Proto byla

provedena datová analýza, z níž získáme vhodně navrženou strukturu ve 3NF, odrážející specifikace zadání a především realitu.

- **Tabulka Projektů**

Obsahuje informace o webovém projektu.

ID, Název, Url, Klíčová_slova, Login, Email, Heslo, Datum_vytvoření, RSS_titulek, RSS_adresa, Firma_Název, Firma_Kontaktní_osoba, Firma_IČ, Firma_DIČ, Firma_Ulice, Firma_Město, Firma_PSC, Firma_Okres, Firma_Telefon, Firma_Fax, Email_Server, Email_Port, Email_SSL, Email_Login, Email_Heslo

- **Tabulky Krátkých a Dlouhých popisů**

Tyto tabulky obsahují možné popisy projektu. Rozdíl je v maximální délce textu.

ID, ID_projektu, Text

- **Tabulka Katalogů**

Zde jsou uloženy základní informace o katalogu, tak jak byly získány ze seznamu katalogů, s přidánými vlastními XML daty s postupem registrace.

ID, Název, Url_adresa, Url_adresa_registrace, S-Rank, PageRank, Typ, Kategorie, Jazyk, XML_data

- **Tabulka Stavů registrace do katalogů**

Obsahuje stav webového projektu k danému katalogu

ID, ID_Katalogu, ID_Projektu, Status

- **Tabulka Kategorii**

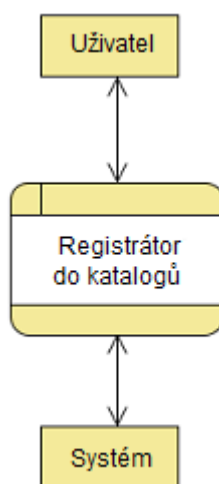
Seznam základních kategorií, jenž se v nějaké podobě nachází na většině katalogů.

ID, Název

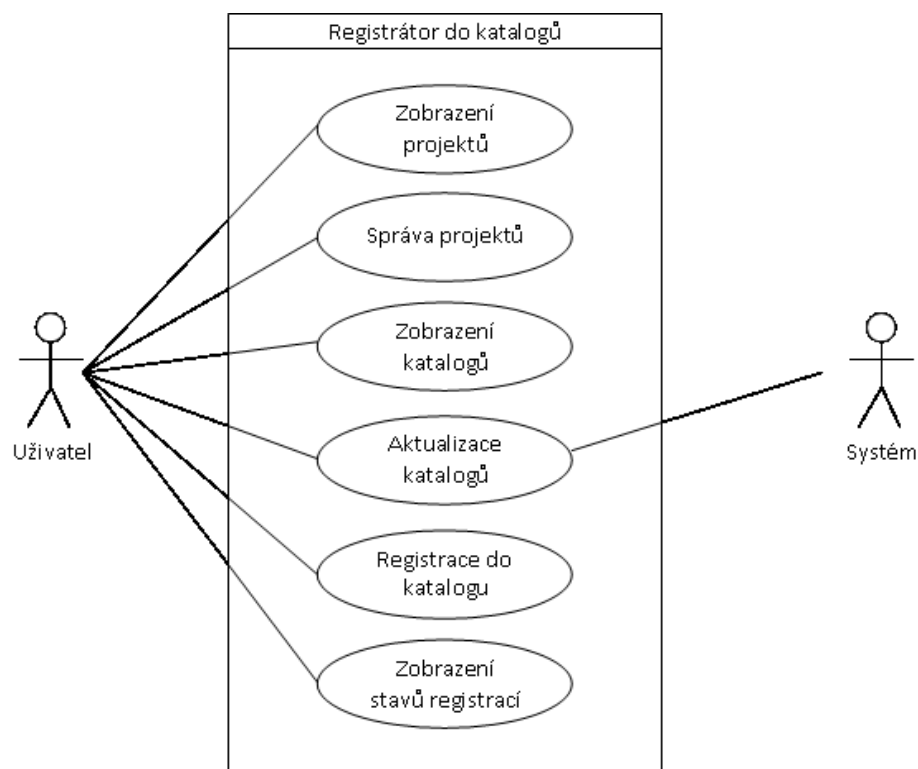
5.3. Funkční analýza

Z obrázku Obrázek 7, na kterém je znázorněno okolí systému, je zřejmé, že se systémem pracuje jak uživatel, tak i samotný systém. Uživatel má k dispozici funkce systému pro výběr a tvorbu projektů, prohlížení seznamu katalogů a samozřejmě registraci do katalogů. Samotný systém pak provádí plánované úlohy jako je aktualizace seznamu katalogů.

Jednotlivé případy užití jsou znázorněny v diagramu případů užití na obrázku Obrázek 8. Vzhledem k tomu, že naplánované úlohy lze spustit i uživatelem, jsou všechny případy užití aktéra „Systém“ i v seznamu aktéra „Uživatel“.



Obrázek 7 - Data flow diagram nulté úrovně - kontextový diagram



Obrázek 8 - Diagram případů užití

- **Aktualizace katalogů**

Aby bylo možné registrovat projekt do katalogu, je potřeba aktualizovat seznam katalogů s daty pro postup registrace. Tato aktualizace může probíhat automaticky (např. po spuštění nebo v určený čas) nebo na pokyn uživatele. Aktualizace probíhá stažením datového souboru ze serveru a následného přečtení a uložení dat do databáze.

- **Registrace do katalogů**

Nejdůležitější funkcí registrátoru je samotná registrace projektu do katalogu. To probíhá v několika základních krocích. Nejdříve se načtou data pro postup registrace z databáze. Z těchto dat zjistíme první krok a přejdeme na stránku, kde tento krok provedeme. Toto opakujeme, dokud nedojdeme k poslednímu kroku.

Jakmile je registrace hotova, je možno provést další kroky k určení, zda registrace byla úspěšná a náš projekt se nachází na stránkách katalogu. V případě, že ani po delší době nedojde k nalezení projektu v katalogu, je potřeba provést opětovnou registraci.

6. Implementace aplikace

Samotná aplikace se skládá ze dvou propojených částí. Samotného registrátoru, který provádí vyplnění dat do formulářů katalogů. A dále z emailového klienta, jenž pomáhá s případy, kdy katalog požaduje potvrzení pomocí návštěvy zaslaného potvrzovacího odkazu.

6.1. Run-time prostředí

Jelikož se jedná o desktopovou aplikaci, bylo na výběr z několika jazyků. Rozhodl jsem se použít jazyk C# z několika důvodů. Jednak se jedná o moderní jazyk, jenž využívá mnoho ze systému Windows, pro který je určen. Dále je to, dle mého názoru, přehledný a dobře čitelný jazyk. Obsahuje také mnoho volně stažitelných balíčků, obsahujících základy řešení pro často používané problémy, jako je přístup k databázi. A v neposlední řadě se jedná o jazyk, se kterým již nějakou dobu pracuji a zvykl jsem si na něj.

Mezi nevýhody tohoto jazyka je již zmíněné provázání se systémem Windows. Bohužel není prakticky možné aplikace napsané v tomto jazyce nějak reálně převést na jiné systémy. Ovšem vzhledem k množství instalací Windows, se nejedná o příliš velký problém.

6.2. Zvolený SŘBD

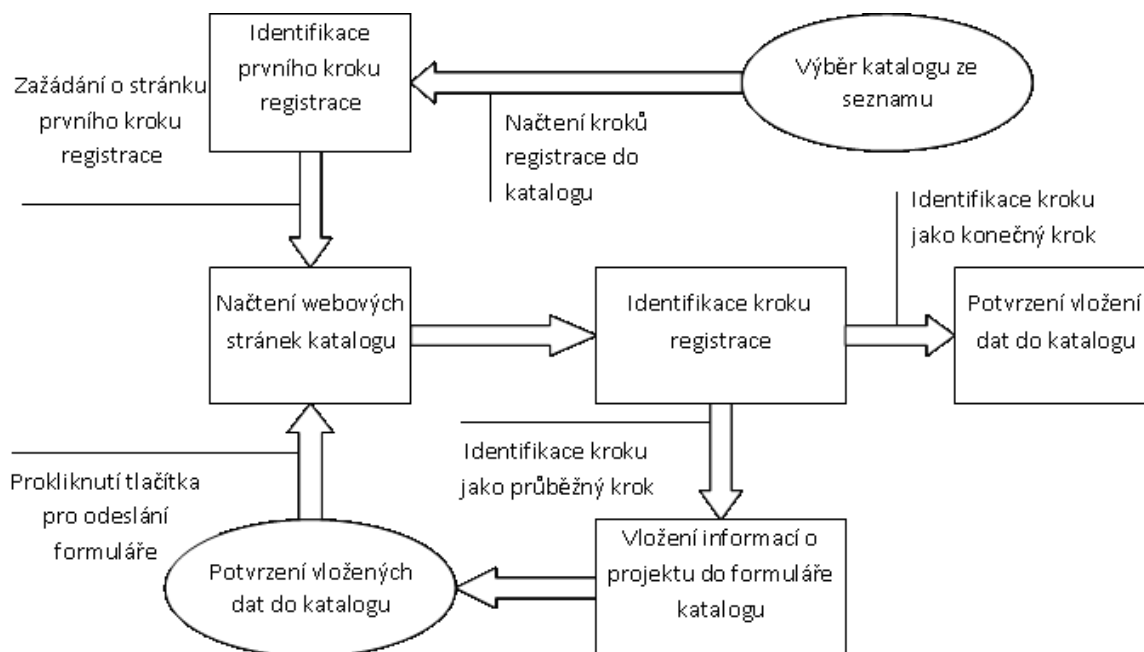
V základu využívá registrátor k uchování dat tabulky Microsoft Access. Je to z důvodu, že pro přístup je potřeba pouze sada Microsoft Office nebo volně přístupné knihovny z této sady, jejichž instalace je přímočará a jednoduchá. Není vyloučeno, že by se v budoucnu přešlo k jinému databázovému systému, např. MS SQL či Oracle. Ovšem v tomto případě by instalace byla více náročná, a to i v případě využití MS SQL Express edice.

6.3. Registrační část

Tato část obsahuje seznam uživatelových projektů s možností jejich editace nebo založení nového projektu, seznam katalogů a způsoby jak vyplnit data určená pro tyto katalogy.

Výhodou použití desktopové aplikace je, že se můžeme vyhnout různým omezením ze strany poskytovatelů katalogů. Jedná se převážně o omezení typu omezený počet registrací z jedné IP adresy v daném časovém úseku. Jelikož uživatel bude v daný čas registrovat pouze jeden projekt, nemůže se stát, že by bylo potřeba v krátkém časovém úseku provést více registrací do jednoho katalogu.

Při psaní kódu jsem se snažil využívat návrhového vzoru MVC. Jedná se o vzor, jenž se nejlépe využívá při psaní webové aplikace, kde lze velmi snadno oddělit prezenční vrstvu od řídicí. V aplikaci je kód prezenční vrstvy oddělen do WinForm třídy tvořící okno aplikace. Kód řídicí vrstvy, obsahující většinu logiky, je umístěna v adresáři Controller. Zde jsou třídy dále děleny podle toho, s kterou částí programu pracují. Nakonec je datová vrstva rozdělena do dvou adresářů. Adresář Data obsahuje kód pro přístup k datům v databázovém souboru, a adresář Model obsahuje logiku pro přístup k datům z jiných zdrojů.

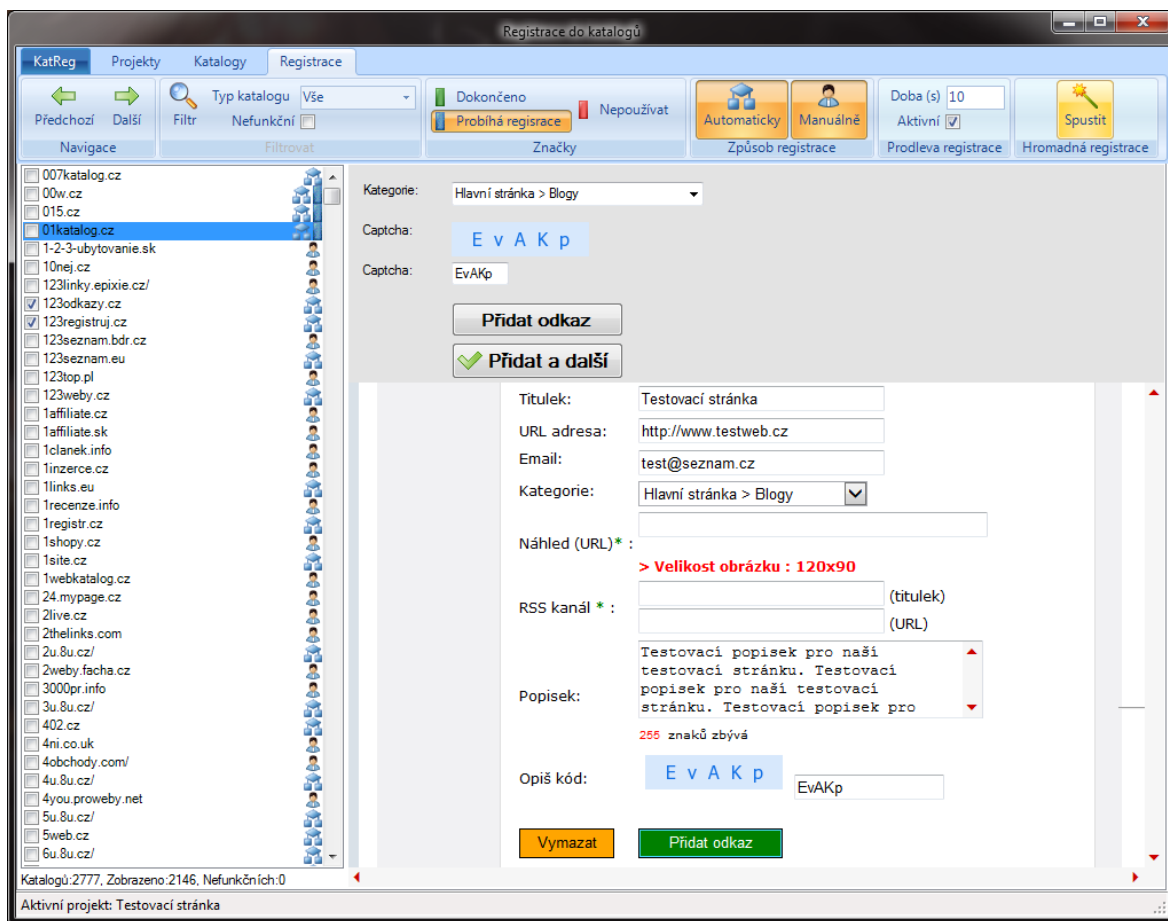


Obrázek 9 – Diagram postupu registrace

Obrázek 10 – Hlavní obrazovka s výběrem projektu

Samotná zpracování registrace probíhá v několika krocích. Úplně prvním krokem je zjištění stránky, kterou je potřeba navštívit jako první. To může být přihlašovací stránka, nebo rovnou stránka pro registraci katalogu. Jakmile aplikace dokončí stahování stránky, zjistí, o který krok v registraci se jedná. Podle toho se použije část XML dat s informacemi pro daný krok.

Ve většině případů se jedná o data identifikující jednotlivé prvky formuláře a jejich následné provázání s daty v databázi. Dále určuje, které tlačítko se má použít pro přesun na další krok v registraci a o který krok v XML datech se jedná. V neposlední řadě obsahuje informace o důležitých dodatečných položkách (data, která jsou příliš specifická pro daný katalog nebo captcha kód).



Obrázek 11 – Automaticky vyplněné prvky v katalogu

6.4. Emailový klient

Jedná se o jednoduchého emailového klienta, přistupující k poštovnímu serveru pomocí protokolu POP3.

Pro použití v aplikaci jsem se rozhodl naimplementovat vlastní POP3 klienta, z důvodů neexistujícího klienta v .NET knihovnách (oproti existujícímu SMTP). Tento klient je omezený pouze na funkce, které jsou pro funkčnost aplikace nezbytné. Jedná se o tyto základní funkce:

- Přihlášení a odhlášení z POP3 serveru
- Seznam zpráv na serveru
- Čtení hlavičky zprávy podle indexu zprávy
- Čtení celé zprávy podle indexu zprávy

I přes omezení jen na potřebné funkce se jedná o téměř kompletního klienta. Mezi nejhlavnější chybějící funkce patří smazání zprávy ze serveru.

```
Received: from smtp1.cpost.cz (smtp1.cpost.cz [193.150.24.99])
    by email-mx16.ng.seznam.cz (Seznam SMTPD UNKNOWN@UNKNOWN) with ESMTP;
    Tue, 03 Jun 2014 19:17:21 +0200 (CEST)
X-IronPort-AV: E=Sophos;i="4.98,967,1392159600";
    d="scan'208,217";a="97260220"
Received: from av5-kasp.centrum.cpost.cz (HELO smtp1.centrum.cpost.cz) ([10.128.31.85])
    by smtp1.cpost.cz with ESMTP; 03 Jun 2014 19:17:21 +0200
Received: from vak5240-6-z5e0 (vak5240-6-z5e0.centrum.cpost.cz [10.160.87.11])
    by smtp1.centrum.cpost.cz (Postfix) with ESMTP id 694CC38049
    for <jirka.kopel@email.cz>; Tue, 3 Jun 2014 19:17:21 +0200 (CEST)
Date: Tue, 3 Jun 2014 19:17:21 +0200 (CEST)
From: =?UTF-8?Q?=C4=8Cesk=C3=A1_po=C5=A1ta?= <ceskaposta@cpost.cz>
To: jirka.kopel@email.cz
Message-ID: <1540915676.254576.1401815841442.JavaMail.esb1_as3@vak5240-6-z5>
Subject: =?UTF-8?Q?Informace_o_Va=C5=A1=C3=AD_z=C3=A1silce_DR9557121179C?=
MIME-Version: 1.0
Content-Type: text/html; charset="utf-8"
Content-Transfer-Encoding: base64
X-Smtpd: UNKNOWN@UNKNOWN
X-Session: 301
X-Country: CZ
```

Zdrojový kód 1 - Příklad hlavičky emailové zprávy

Pro čtení pouze hlavičky zprávy je použit příkaz **TOP x y**, který vrací **y** řádků těla zprávy **x**. Tato funkce je použita ke stažení seznamu zpráv a rozlišení zda zprávu je potřeba číst celou nebo zda zpráva již byla přečtena.

Při čtení celé zprávy je nejdůležitějším krokem zjištění, jaký typ těla zpráva obsahuje. Základní typy těla zprávy jsou *čistý text* (text/plain), *HTML text* (text/html) a *smíšený obsah* (multipart/mixed). V dnešní době asi nejčastěji používaný typ je smíšený obsah. V takovémto těle se většinou nachází text zprávy jako čistý text a zároveň i jako HTML text. Dále obsahuje veškeré přílohy zprávy. V mém klientu tyto jednotlivé části těla zprávy rozdělují. Pokud se jedná o textovou část (čistý i HTML text) tak text používám k nalezení potřebných odkazů. Zbylé části těla jako jsou obrázky pro HTML zobrazení a různé přílohy stahuji z důvodu, že předem nemůžu vědět, jak velkou část zprávy budu potřebovat.

Dalším důležitým krokem je detekce a překlad syntaxe „Encoded-word“. Tato syntaxe umožňuje pro převod textu znaků jiné znakové sady do ASCII a je používána převážně v hlavičce zprávy, která musí obsahovat pouze ASCII znaky. Nejčastěji

můžeme vidět tuto syntaxi použitou v části hlavičky obsahující předmět zprávy nebo odesílatele.

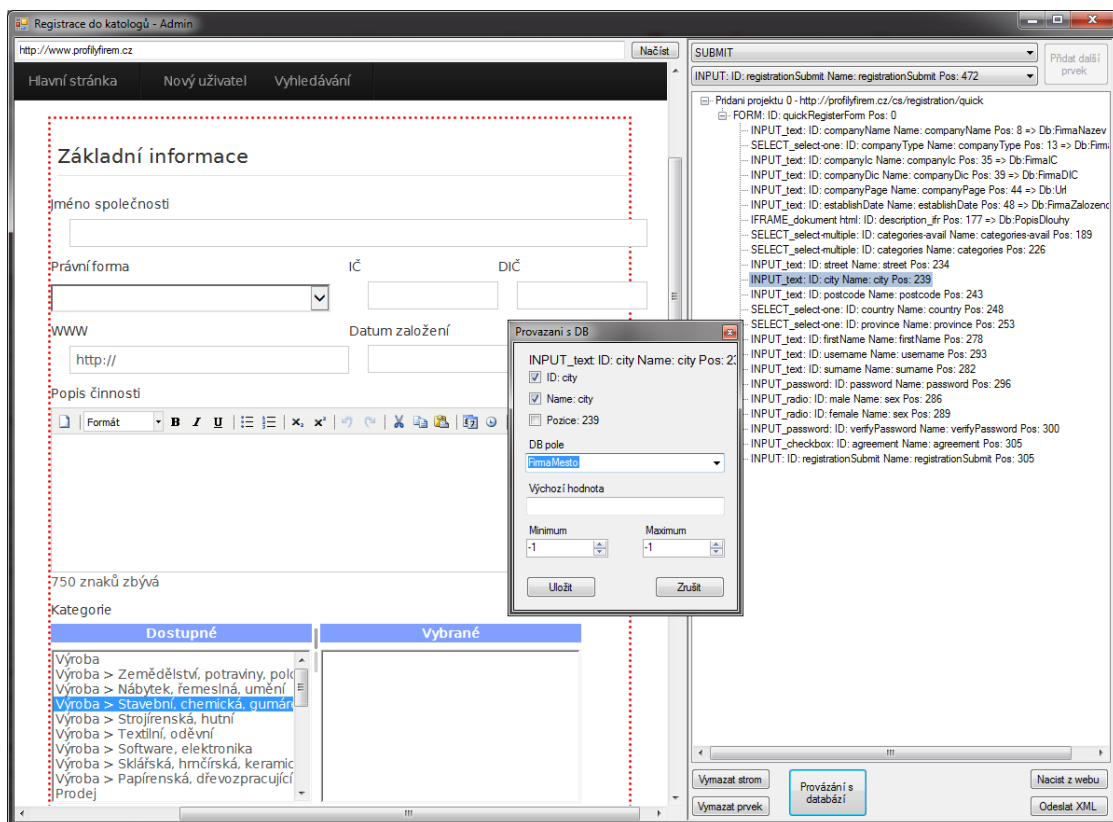
6.5. Administrační aplikace

Tato aplikace je určena pro vytváření a aktualizace datových XML souborů používaných pro rozhodování jednotlivých kroků registrace. Jedná se o vcelku jednoduchou aplikaci, která obsahuje panel pro zobrazení webové stránky a druhý panel, jenž zobrazuje data pro již zadané kroky registrace.

Jednotlivé kroky pro vytvoření registračních dat jsou jednoduché. Nejdříve navštívíme stránku katalogu. V něm navštívíme stránku, která bude naším prvním krokem v registraci (většinou se jedná o registrační stránku nebo o stránku pro přihlášení). Zde vybereme ze seznamu formulářů, který webový formulář obsahuje prvky důležité pro daný krok. Mnohdy tyto stránky obsahují další formuláře, např. pro vyhledávání v katalogu nebo pro přihlášení do katalogu.

Dalším krokem po vybrání formuláře je vybrání jednotlivých prvků formuláře a jejich následné provázání na data registrovaného projektu. Důležitá část je vybrání tlačítka formuláře pro pokračování v registraci a provázání na odpovídající akci. Toto provázání je nezbytné pro správnou funkčnost registrační aplikace a umožňuje provést aplikační stisknutí tohoto tlačítka.

Mezi další nepřímé provázání na data projektu patří identifikace záchytného obrázku nebo další neobvyklé otázky.



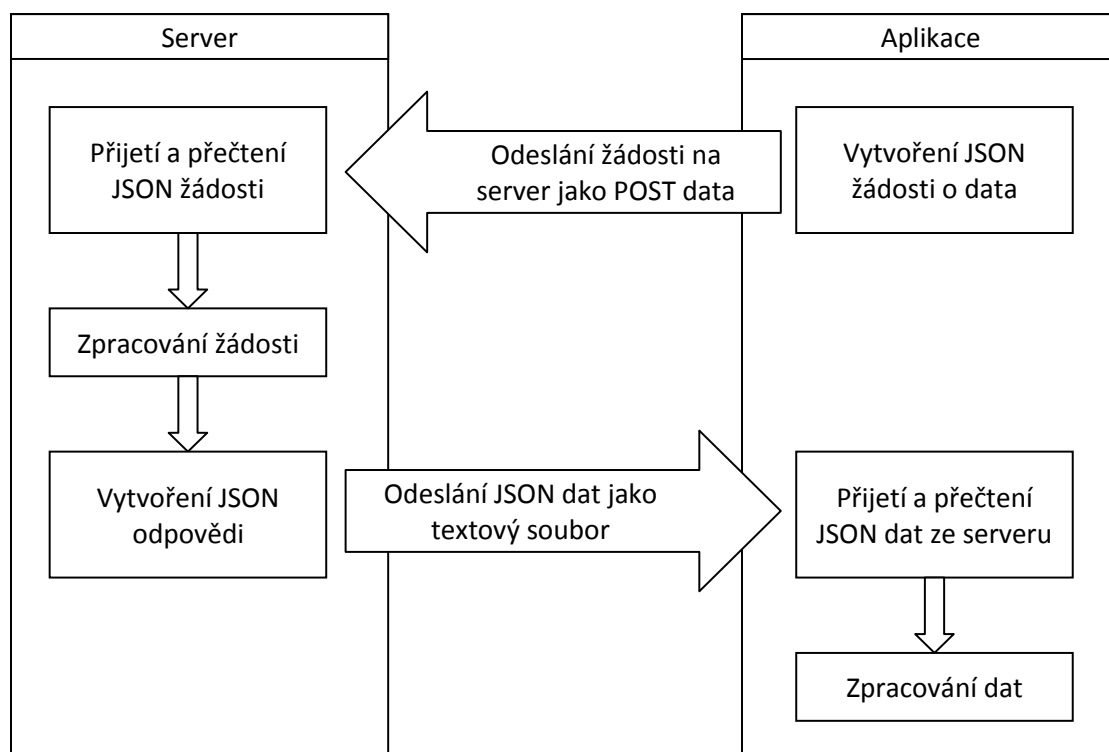
Obrázek 12 – Administrační program ve stádiu provazování formulářového pole s databází

6.6. Aktualizace aplikace ze serveru

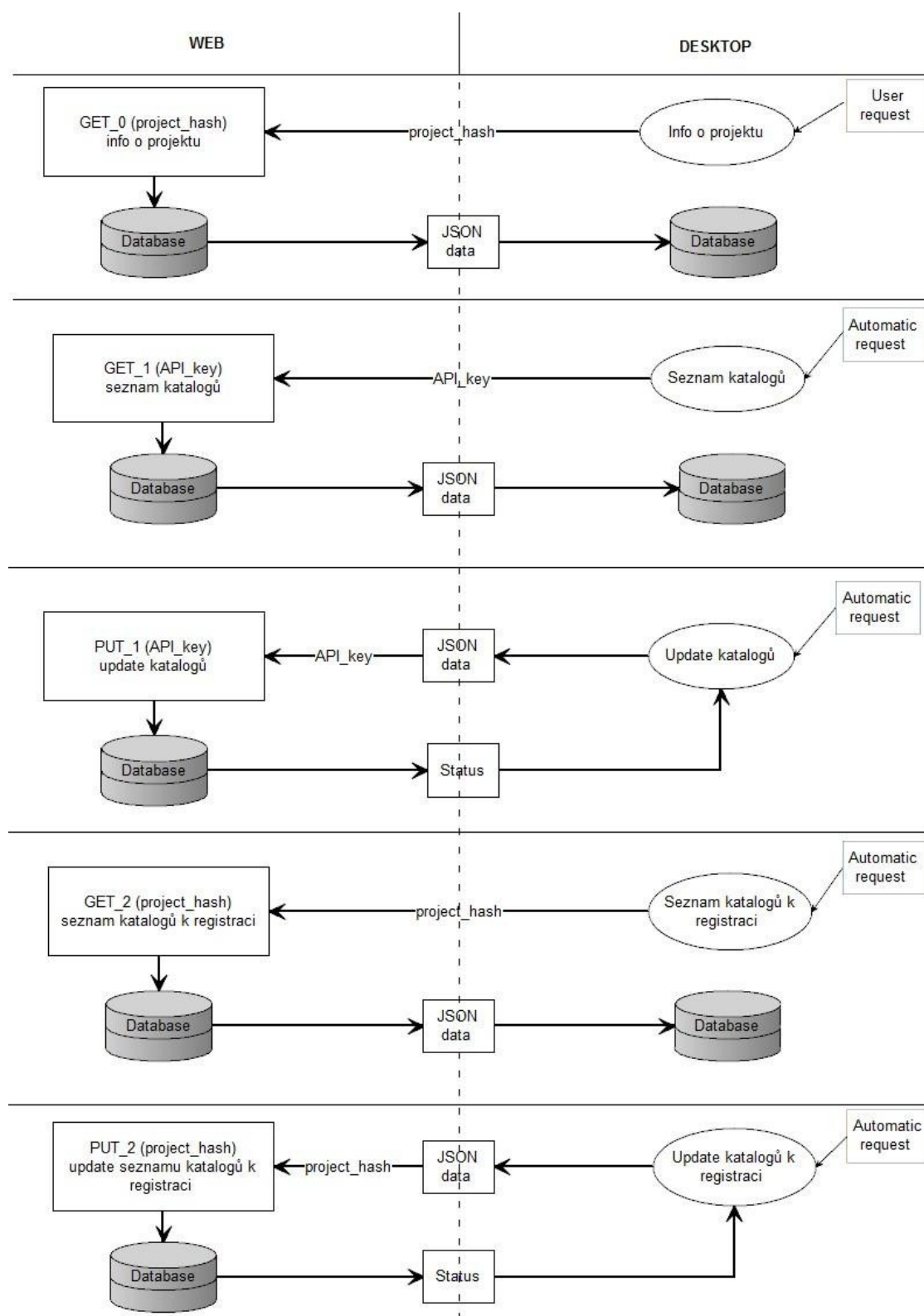
Toto API bude sloužit pro komunikaci a výměnu dat mezi webovou a desktopovou aplikací, která provádí automatickou registraci do katalogů. Webová část poskytuje databázi pro uložení informací o uživatelských projektech, celkovém seznamu katalogů a seznamu katalogů k registraci.

Informace o projektu budou poskytnuty po ověření HASHe projektu, který uživatel vyplní v desktopové aplikaci. Seznam katalogů a jejich nastavení se bude stahovat automaticky a ověřovat pomocí API klíče. Webová databáze katalogů bude zpětně aktualizována desktopovou aplikací. Aplikace si bude stahovat seznam katalogů, na kterých chce uživatel provést registraci. Desktop bude vracet seznam katalogů, na kterých provedl registraci, spolu s jejich stavem, v jakém se nachází.

Data jsou přenášena ve formátu JSON.



Obrázek 13- Postup aktualizace dat ze serveru



Obrázek 14 – API přenosu dat mezi serverem a aplikací

7. Využité technologie a nástroje

7.1. Microsoft Visual Studio 2012

Visual Studio je program určený ke zjednodušení psaní kódu v několika jazycích. Mezi hlavní jazyky patří Visual Basic.NET a C#. Další velmi používaný jazyk je LINQ. Jedná se o jazyk usnadňující dotazování a práci s daty. Lze ho propojit téměř s jakýmkoliv zdrojem dat, např. s MS SQL, MySQL nebo i XML

- **Výhody práce ve Visual Studio**

Visual Studio obsahuje řadu nástrojů, které nám velice zjednoduší naši práci. Jedním z nich je design mód, který umožňuje přímé vkládání objektů z Toolboxu na formulář. Výsledný zdrojový kód, nutný pro zobrazení objektu, se nám generuje automaticky.

Dalším velmi používaným nástrojem je IntelliSense, který pomáhá omezit počet chyb a urychluje psaní kódu. Při psaní se nám automaticky generuje nabídka všech možných příkazů, proměnných a definic. Dále Visual Studio obsahuje možnosti ladění aplikací, různé průvodce pro tvorbu databází atd.

7.2. LINQ

Jedná se o dotazovací jazyk který byl představen spolu s .NET Frameworkem 3.5. LINQ je navrhnut jako obecný nástroj umožňující manipulaci s různými daty. Mezi výhody LINQ patří objektový přístup k datům.

Mezi základní LINQ poskytovatele patří:

- **LINQ to Objects** - Slouží pro dotazování nad daty uloženy v paměti (pole a třídy implementující rozhraní `IEnumerable<T>`)
- **LINQ to SQL** - Slouží pro dotazování nad databázemi využívající rozhraní MS SQL.
- **LINQ to XML** - Slouží pro práci s daty uložených v XML souborech.
- **LINQ to DataSets** - Slouží pro dotazování nad obecnými databázemi.

Dají se sehnat i LINQ poskytovatelé třetích stran, které umožňují přístup k množství dalších zdrojů dat. Mezi zajímavější patří například: LINQ to Google, LINQ to CSV, LINQ to Twitter, LINQ to Wikipedia. Aplikace využívá LINQ to Db, jenž je určen pro přístup k celé řadě databázových zdrojů.

I přestože primárním přínosem LINQ je maximální efektivita při dotazování nad daty, u většiny implementací LINQ to u dotazování nekončí a kromě toho je možné data nejen efektivně dotazovat, ale i také modifikovat. Takže například v případě LINQ to SQL nám do rukou přichází nástroj, který nejen, že umožní provádět dotazy nad databází SQL server jednoduše a typově ze C# 3.0, VB 9 nebo jiného jazyku, ale díky možnosti data i modifikovat se z této implementace LINQ stává skvěle použitelný objektově-relační mapper. [5]

7.3. Json

JSON je způsob zápisu dat. Je člověkem i strojově jednoduše čitelný a zapisovatelný. Jedná se o textový, na jazyce nezávislý formát. Využívá konvence známé z rodiny jazyků C (C, C++, C#, Java, JavaScript, Perl, Python a další).

Mezi nedostatky JSON patří to, že neumožňuje definovat znakovou sadu přenášeného obsahu, optimalizace pro přenos binárních dat. Tyto nedostatky platí ale pro některé (slabší) implementace. Nealfabetické znaky v řetězcích a binární data JSON jsou escapovány zpětným lomítkem, za kterým následuje buď jeden z běžně používaných znaků (např. \n pro nový řádek, \t pro tabulátor, \\ pro samotné zpětné lomítko) nebo \u indikující znak z Unicode (UTF-16), za nímž následují čtyři hexadecimální číslice. [6]

7.4. Podmínky pro funkčnost Aplikace

Pro funkčnost aplikace je zapotřebí splnění několika podmínek na nainstalovaný software. Základním předpokladem je přítomnost *Microsoft .NET Framework 4.0* nebo novější. Dále je potřebná přítomnost *Microsoft Office Access Database Engine 2007*, jenž je součástí instalace sady *Microsoft Office 2007* nebo lze v případě nutnosti nainstalovat samostatně.

8. Závěr

Cílem této práce byla analýza a implementace aplikace, která by řešila automatizaci procesu registrace do internetových katalogů a vyhledávačů se zaměřením na problematiku automatizovaného a polo-automatizovaného registrování do internetových katalogů. Vytvořil jsem tedy aplikaci s obecným algoritmem, který provádí polo či plně automatickou registraci do katalogu na základě snadno aktualizovaných datových souborů. Dále administrační aplikaci, sloužící pro vytváření nové verze datového souboru a jejich následné nahrání na server.

Plná automatizace postupu registrace do všech katalogů je technicky velmi složitá. Hlavním důvodem jsou různé druhy zabezpečení. Z nich nejobtížnější jsou Záchytné obrázky, které existují v mnoha formách. Z tohoto důvodu se bez alespoň minimálního zásahu uživatele neobejdeme.

Použití datových souborů stahovaných přes aktualizací server umožňuje snadnou aktualizaci a rozšíření seznamu katalogů. Toto má za výsledek možnost dlouhodobé životnosti aplikace s pouhou potřebou udržování aktuálních datových souborů.

9. Citovaná literatura

1. **GRAPPONE, Jennifer a COUZIN, Gradiva.** *SEO: search engine optimization : ovládněte SEO a získejte výhodu před konkurencí : optimalizujte své webové stránky pro vyhledávací servery : přiveďte na své stránky zákazníky dříve, než to udělá konkurence.* [překl.] Roman Skřivánek a Dana Balaštíková. Brno : Zoner Press, 2007. ISBN 978-80-86815-85-5.
2. **Linhart, Jan a Kubíček, Michal.** *333 tipů a triků pro SEO.* Brno : Computer Press, 2010. ISBN: 978-80-251-2468-0.
3. **Kubíček, Michal.** *Velký průvodce SEO: Jak dosáhnout nejlepších pozic ve vyhledávačích.* Brno : Computer Press, 2008.
4. CAPTCHA – Wikipedie. *Wikipedie, otevřená encyklopedie.* [Online] 7. Srpen 2013. [Citace: 30. Červen 2014.] <http://cs.wikipedia.org/wiki/CAPTCHA>.
5. HTTP referrer – Wikipedie. *Wikipedie, otevřená encyklopedie.* [Online] 5. Duben 2013. [Citace: 30. Červen 2014.] http://cs.wikipedia.org/wiki/HTTP_referrer.
6. Informační systém - Wikipedie. *Wikipedie, otevřená encyklopedie.* [Online] 23. 11 2012. [Citace: 27. 11 1012.] http://cs.wikipedia.org/wiki/Informa%C4%8Dn%C3%AD_syst%C3%A9m.
7. **Puš, Petr.** Úvod do LINQ | Vyvojar.cz. *Vyvojar.cz - Vývojáři sobě.* [Online] 24. Leden 2008. [Citace: 28. Červen 2014.] <http://www.vyvojar.cz/Articles/563-uvod-do-linq.aspx>.
8. JavaScript Object Notation – Wikipedie. *Wikipedie, otevřená encyklopedie.* [Online] 19. Říjen 2013. [Citace: 29. Červen 2014.] <http://cs.wikipedia.org/wiki/JSON>.